Spatial Distribution and Diverse Metabolic Functions of Lignocellulose-Degrading Uncultured Bacteria as Revealed by Genome-Centric Metagenomics

Kougias, Panagiotis ; Campanaro, Stefano; Treu, Laura; Tsapekos, Panagiotis; Armani, Andrea; Angelidaki, Irini

Published in:
Applied and Environmental Microbiology

Link to article, DOI:
10.1128/AEM.01244-18

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

BIOTECHNOLOGY

Spatial Distribution and Diverse Metabolic Functions of Lignocellulose-Degrading Uncultured Bacteria as Revealed by Genome-Centric Metagenomics

Panagiotis G. Kougias,[a] Stefano Campanaro,[b] Laura Treu,[a] Panagiotis Tsapekos,[a] Andrea Armani,[c] Irini Angelidaki[a]

[a]Department of Environmental Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
[b]Department of Biology, University of Padua, Padua, Italy
[c]Department of Biomedical Science, University of Padua, Padua, Italy

ABSTRACT The mechanisms by which specific anaerobic microorganisms remain firmly attached to lignocellulosic material, allowing them to efficiently decompose organic matter, have yet to be elucidated. To circumvent this issue, microbiomes collected from anaerobic digesters treating pig manure and meadow grass were fractionated to separate the planktonic microbes from those adhered to lignocellulosic substrate. Assembly of shotgun reads, followed by a binning process, recovered 151 population genomes, 80 out of which were completely new and were not previously deposited in any database. Genome coverage allowed the identification of microbial spatial distribution in the engineered ecosystem. Moreover, a composite bioinformatic analysis using multiple databases for functional annotation revealed that uncultured members of the *Bacteroidetes* and *Firmicutes* follow diverse metabolic strategies for polysaccharide degradation. The structure of cellulosome in *Firmicutes* species can differ depending on the number and functional roles of carbohydrate-binding modules. In contrast, members of the *Bacteroidetes* are able to adhere to and degrade lignocellulose due to the presence of multiple carbohydrate-binding family 6 modules in beta-xylosidase and endoglucanase proteins or S-layer homology modules in unknown proteins. This study combines the concept of variability in spatial distribution with genome-centric metagenomics, allowing a functional and taxonomical exploration of the biogas microbiome.

IMPORTANCE This work contributes new knowledge about lignocellulose degradation in engineered ecosystems. Specifically, the combination of the spatial distribution of uncultured microbes with genome-centric metagenomics provides novel insights into the metabolic properties of planktonic and firmly attached to plant biomass bacteria. Moreover, the knowledge obtained in this study enabled us to understand the diverse metabolic strategies for polysaccharide degradation in different species of *Bacteroidetes* and *Clostridiales*. Even though structural elements of cellulosome were restricted to *Clostridiales* species, our study identified a putative mechanism in *Bacteroidetes* species for biomass decomposition, which is based on a gene cluster responsible for cellulose degradation, disaccharide cleavage to glucose, and transport to cytoplasm.

KEYWORDS archaea, anaerobic digestion, lignocellulose, metagenomics, methane, microbial ecology, uncultured microbes, metabolism

The mechanisms of several microbial processes remain unclear due to the fact that only a small fraction of microorganisms that exists in nature can be cultivated (1). Especially in complex microcosms, such as the one responsible for anaerobic digestion of organic matter, understanding of the microbial dynamicity is additionally hampered by the presence of syntrophic interactions between members of the community. In past

Received 22 May 2018 **Accepted** 6 July 2018
Accepted manuscript posted online 13 July 2018
Citation Kougias PG, Campanaro S, Treu L, Tsapekos P, Armani A, Angelidaki I. 2018. Spatial distribution and diverse metabolic functions of lignocellulose-degrading uncultured bacteria as revealed by genome-centric metagenomics. Appl Environ Microbiol 84:e01244-18. https://doi.org/10.1128/AEM.01244-18.
Editor Shuang-Jiang Liu, Chinese Academy of Sciences

Address correspondence to Laura Treu, latr@env.dtu.dk.
P.G.K. and S.C. contributed equally to this work.

years, different molecular techniques, ranging from 16S rRNA gene sequencing to metagenomics, were proposed to overcome these obstacles and to explore the functional principles of the unculturable microbial majority.

In comparison to 16S rRNA gene-based studies, metagenomic investigations using shotgun sequencing became more attractive and were applied to anaerobic digestion systems to obtain relevant insights into the metabolic properties of the whole microbiome (2). However, the reads generated by high-throughput sequencing are short, error prone, and contain limited signal for homology searches; thus, shotgun assemblies rendered the obtained results more reliable (3, 4). Moreover, all of these gene-centric metagenomic studies consist of inventories of individual annotated genes (5) and do not classify the genes to single population genomes (PG). Recent advances in bioinformatics and the introduction of automated strategies for binning process, such as Clustering cONtigs with COverage and ComposiTion (CONCOCT), GroopM, or MetaBAT (6–8), shift the perspective of metagenomics from gene-centric to genome-centric. Quite recently, this novel approach was applied to the anaerobic digestion microbiome, leading to the extraction of hundreds of PGs, which were assigned to the four steps of the process, i.e., hydrolysis, acidogenesis, acetogenesis, and methanogenesis (9–11). Genome-centric metagenomics have the undoubted benefit of revealing the functional properties of individual genomes, leading to a more detailed comprehension of the microbial interactions occurring in the microbiome (12). Additionally, the identification of the sequenced genomes can be combined with metatranscriptomic data, which provides further information regarding the metabolic basis for adaptation of particular phylotypes (5) and allows the monitoring of gene expression changes at the single-species level (13).

The microbiomes associated with the conversion of lignocellulose-derived sugars into biofuels and bioproducts are gaining particular attention in the wake of climate change. Lignocellulosic particles are typically present in biogas digesters, as they can be a source of influent feedstock (e.g., grass, maize etc.) or constitute a large fraction of other feeding residues (e.g., fibers in manure). It is well known that the hydrolysis step of the anaerobic lignocellulose digestion process is mainly facilitated by some anaerobic bacteria that are able to produce a multienzyme complex, called cellulosome, and by aerobic bacteria and fungi that typically produce free, noncomplexed individual enzymes (14, 15). Moreover, the recalcitrant nature of the lignocellulosic substrate renders the hydrolysis process as a rate-limiting step in the anaerobic digestion food chain (16). Therefore, a more detailed knowledge of the firmly attached species and their interactions with the planktonic microbiome is essential for the development of strategies that will accelerate the hydrolysis rate. In turn, the overall efficiency of biotechnological processes (e.g., production of biofuels and bio-based platform chemicals) will be significantly improved.

Previous studies targeting the identification of the microbial species firmly attached to the plant material were based only on 16S rRNA gene-based terminal restriction fragment length polymorphism (17) or fluorescent *in situ* hybridization (FISH; 18). Recently, metagenomic binning analysis and metatranscriptomics were performed to identify species involved in cellulose degradation (10, 19). Nevertheless, these high-throughput sequencing methods did not identify the spatial distribution of the microbial species. Moreover, detailed genomic analyses of polysaccharide binding and degradation pathways were focused on few microbial species, with *Clostridium thermocellum* recognized as the most interesting model (20). However, limited information is available for uncultured species of the phylum *Bacteroidetes* (21), which are known to play a relevant role in proficient plant biomass-degrading ecosystems and in anaerobic digestion sludge (19, 22). The importance of *Bacteroidetes* species in the thermophilic cellulose methanation process was previously underlined (19), and therefore the sequencing of these genomes is extremely important to shed light on their metabolic potential.

In the present study, a genome-centric approach was applied to elucidate the spatial distribution of lignocellulose degradation mechanisms in anaerobic digestion

process. This was done because it was hypothesized that the localization of microorganisms should be mainly determined by their functional roles, and thus the microbes that are firmly attached to the plant material should be responsible for the degradation of lignocellulosic biomass. Samples obtained from biogas reactors (i.e., firmly attached to grass particles and planktonic phases from codigestion systems and independent manure samples from monodigestion systems) were subjected to high-throughput shotgun sequencing (Illumina). A subsequent *de novo* assembly and binning process led to the extraction of 151 population genomes. Bioinformatics functional analysis performed on the recovered draft genomes allowed a detailed characterization of the functional structure of the microbiome, and, more importantly, clarified the strategies followed by different species to bind to the solid fraction of plant material and to perform polysaccharide hydrolysis.

RESULTS

A targeted fractionation strategy was performed to analyze the spatial localization of the microbiome involved in lignocellulose degradation in biogas reactors. More specifically, nine samples were collected (i.e., three biological replicates per each treatment), representing the microbial consortia residing in the liquid grass fraction (LG), or firmly attached to grass biomass (FG), or in an anaerobic system treating pig manure without lignocellulose (PM). Both reactors were operating at stable conditions during the sampling and were able to efficiently produce methane at more than 84% of the theoretical potential (Supplemental File S1).

Approximately 120 million paired-end reads were coassembled to obtain a reliable representation of the genomes present in the microbial communities. This process resulted in the generation of 483,935 scaffolds ranging in size from 500 to 371,582 bp (N_{50}, 2,426 bp; total length, 775,466,133 bp). Gene prediction performed on the scaffolds led to the identification of 1,102,319 protein-encoding sequences whose functions were investigated using five software packages based on different databases (i.e., Kyoto Encyclopedia of Genes and Genomes [KEGG], Clusters of Orthologous Groups [COG], DataBase for automated Carbohydrate-active enzyme ANnotation [dbCAN], SEED, Pfam). The combination of diverse annotation strategies provided a detailed overview of the functional properties of the metagenome.

The applied binning process allowed the extraction of 151 PGs from the assembly (Fig. 1) with an estimated completeness ranging between ~11% and 100% (average 73%) (Fig. 1 and Supplemental Data Set S1). The superlative outcome of the assembly and binning processes was evidenced by the 120 PGs and the 94 PGs having an estimated completeness of more than 50% and 70%, respectively. Moreover, 45 PGs fulfilled the stringent criteria for defining "high-quality draft" genomes (i.e., >90% complete with less than 5% contamination) according to recently published standard metrics (23). PGs with less than 50% completeness were only used to define the taxonomy and to check species abundance in the three different fractions. Thus, the functional characteristics derived from their gene content are not included in the Discussion.

The taxonomic structure of the biogas microbiome was dominated by members of the *Firmicutes* (101 PGs), followed by the *Bacteroidetes* (11 PGs), *Gammaproteobacteria* (8 PGs), *Synergistia* (7 PGs), and some other less frequently found phyla (Fig. 1 and Supplemental Data Set S1). Archaea were represented by 5 PGs; four of these belonging to *Methanomicrobia* and one to *Methanobacteria* (Fig. 1). All of the extracted PGs were present in the three fractions (i.e., FG, LG, and PM) in different coverage levels, except for one methanogen (*Euryarchaeota* sp. DTU008) that was absent from FG samples.

Analysis of PG coverage in the 9 examined samples revealed that most of the reconstructed genomes ($n = 64$) were enriched more than 2-fold in samples collected from the liquid fraction (LG), while only 25 species were found to increase more than 2-fold in the samples collected from the grass biomass (FG) (Fig. 1 and 2). The remaining 62 PGs did not present any significant change in any of these two fractions

PG name	PG ID	Compl.	Redud.	Fraction
Acholeplasmatales sp. DTU056	Fi12_BG			
Alcaligenaceae sp. DTU414	Pr02_BG			
Anaerobaculum sp. DTU043	Sy02_BG			
Arcobacter sp. DTU416	Pr08_BG			
Arcobacter sp. DTU420	Pr13_BG			
Bacilli sp. DTU057	Fi41_BG			
Bacteroidales sp. DTU141	Ba09_BG			
Bacteroidales sp. DTU338	Ba10_BG			
Bacteroidales sp. DTU360	Ba11_BG			
Bacteroidetes sp. DTU134	Ba02_BG			
Bacteroidetes sp. DTU139	Ba06_BG			
Bacteroidetes sp. DTU149	Ba04_BG			
Campylobacteraceae sp. DTU422	Pr16_BG			
Campylobacterales sp. DTU103	Pr05_BG			
Candidatus Desulforudis sp. DTU372	Fi13_BG			
Cardiobacteriaceae sp. DTU421	Pr15_BG			
Chloroflexi sp. DTU161	Ch02_BG			
Chloroflexi sp. DTU361	Ch01_BG			
Chloroflexi sp. DTU362	Ch03_BG			
Clostridia sp. DTU014	Fi10_BG			
Clostridia sp. DTU011	Fi03_BG			
Clostridia sp. DTU025	Fi38_BG			
Clostridia sp. DTU062	Fi19_BG			
Clostridia sp. DTU065	Fi33_BG			
Clostridia sp. DTU077	Fi36_BG			
Clostridia sp. DTU080	Fi61_BG			
Clostridia sp. DTU084	Fi64_BG			
Clostridia sp. DTU095	Fi70_BG			
Clostridia sp. DTU183	Fi24_BG			
Clostridia sp. DTU196	Fi04_BG			
Clostridiaceae sp. DTU059	Fi20_BG			
Clostridiaceae sp. DTU079	Fi53_BG			
Clostridiaceae sp. DTU392	Fi68_BG			
Clostridiaceae sp. DTU399	Fi79_BG			
Clostridiaceae sp. DTU402	Fi84_BG			
Clostridiaceae sp. DTU413	Fi99_BG			
Clostridiale sp. DTU023	Fi82_BG			
Clostridiales Family XIII. Incertae Sedis sp. DTU377	Fi30_BG			
Clostridiales sp. DTU128	Fi26_BG			
Clostridiales sp. DTU012	Fi07_BG			
Clostridiales sp. DTU013	Fi05_BG			
Clostridiales sp. DTU020	Fi40_BG			
Clostridiales sp. DTU024	Fi59_BG			
Clostridiales sp. DTU031	Fi39_BG			
Clostridiales sp. DTU033	Fi46_BG			
Clostridiales sp. DTU035	Fi32_BG			
Clostridiales sp. DTU036	Fi18_BG			
Clostridiales sp. DTU055	Fi09_BG			
Clostridiales sp. DTU066	Fi29_BG			
Clostridiales sp. DTU069	Fi48_BG			
Clostridiales sp. DTU070	Fi06_BG			
Clostridiales sp. DTU074	Fi58_BG			
Clostridiales sp. DTU076	Fi17_BG			
Clostridiales sp. DTU078	Fi42_BG			
Clostridiales sp. DTU081	Fi47_BG			
Clostridiales sp. DTU083	Fi49_BG			
Clostridiales sp. DTU092	Fi43_BG			
Clostridiales sp. DTU093	Fi77_BG			
Clostridiales sp. DTU099	Fi75_BG			
Clostridiales sp. DTU126	Fi86_BG			
Clostridiales sp. DTU228	Fi76_BG			
Clostridiales sp. DTU229	Fi65_BG			
Clostridiales sp. DTU234	Fi54_BG			
Clostridiales sp. DTU368	Fi102_BG			
Clostridiales sp. DTU370	Fi104_BG			
Clostridiales sp. DTU371	Fi107_BG			
Clostridiales sp. DTU381	Fi45_BG			
Clostridiales sp. DTU386	Fi57_BG			
Clostridiales sp. DTU393	Fi69_BG			
Clostridiales sp. DTU394	Fi71_BG			
Clostridiales sp. DTU396	Fi73_BG			
Clostridiales sp. DTU397	Fi74_BG			
Clostridiales sp. DTU405	Fi89_BG			
Clostridiales sp. DTU407	Fi91_BG			
Clostridiales sp. DTU408	Fi93_BG			

PG name	PG ID	Compl.	Redud.	Fraction
Clostridiales sp. DTU410	Fi95_BG			
Clostridiales sp. DTU412	Fi97_BG			
Clostridiales Family XI sp. DTU367	Fi101_BG			
Clostridiales Family XI sp. DTU374	Fi23_BG			
Clostridiales Family XI sp. DTU379	Fi35_BG			
Clostridiales Family XI sp. DTU380	Fi44_BG			
Clostridiales Family XI sp. DTU383	Fi51_BG			
Clostridiales Family XI sp. DTU384	Fi52_BG			
Clostridiales Family XI sp. DTU388	Fi62_BG			
Clostridiales Family XI sp. DTU390	Fi66_BG			
Clostridiales Family XI sp. DTU400	Fi80_BG			
Clostridiales Family XI Incertae Sedis sp. DTU349	Fi98_BG			
Clostridiales Family XI Incertae Sedis sp. DTU350	Fi14_BG			
Dysgonomonas sp. DTU356	Ba03_BG			
Dysgonomonas sp. DTU357	Ba05_BG			
Dysgonomonas sp. DTU358	Ba07_BG			
Dysgonomonas sp. DTU359	Ba08_BG			
Erysipelothrix sp. DTU091	Fi85_BG			
Erysipelothrix sp. DTU403	Fi87_BG			
Eubacteriaceae sp. DTU369	Fi103_BG			
Eubacteriaceae sp. DTU376	Fi27_BG			
Eubacteriaceae sp. DTU395	Fi72_BG			
Euryarchaeota sp. DTU008	Eu04_BG			
Firmicutes sp. DTU176	Fi11_BG			
Firmicutes sp. DTU226	Fi15_BG			
Firmicutes sp. DTU387	Fi60_BG			
Firmicutes sp. DTU389	Fi63_BG			
Firmicutes sp. DTU398	Fi78_BG			
Firmicutes sp. DTU411	Fi96_BG			
Gammaproteobacteria sp. DTU037	Pr01_BG			
Gammaproteobacteria sp. DTU040	Pr11_BG			
Gammaproteobacteria sp. DTU252	Pr07_BG			
Gammaproteobacteria sp. DTU415	Pr06_BG			
Gammaproteobacteria sp. DTU418	Pr10_BG			
Gammaproteobacteria sp. DTU419	Pr12_BG			
Halothermothrix sp. DTU029	Fi08_BG			
Halothermothrix sp. DTU391	Fi67_BG			
Methanoculleus sp. DTU006	Eu01_BG			
Methanoculleus sp. DTU007	Eu02_BG			
Methanomicrobiales sp. DTU363	Eu05_BG			
Methanosarcina sp. DTU364	Eu06_BG			
Methanothermobacter sp. DTU051	Eu03_BG			
Mollicutes sp. DTU373	Fi16_BG			
Moorella sp. DTU313	Fi31_BG			
Pelotomaculum sp. DTU098	Fi28_BG			
Pelotomaculum sp. DTU382	Fi50_BG			
Pelotomaculum sp. DTU401	Fi81_BG			
Peptococcaceae sp. DTU027	Fi55_BG			
Peptococcaceae sp. DTU404	Fi88_BG			
Proteobacteria sp. DTU431	UC01_BG			
Pseudomonadaceae sp. DTU417	Pr09_BG			
Rikenellaceae sp. DTU001	Ba01_BG			
Ruminococcaceae sp. DTU365	Fi01_BG			
Ruminococcaceae sp. DTU409	Fi94_BG			
Spirochaeta sp. DTU042	Sp06_BG			
Spirochaetales sp. DTU423	Sp01_BG			
Spirochaetales sp. DTU424	Sp03_BG			
Synergistaceae sp. DTU044	Sy01_BG			
Synergistaceae sp. DTU109	Sy06_BG			
Synergistaceae sp. DTU427	Sy03_BG			
Synergistaceae sp. DTU428	Sy04_BG			
Synergistaceae sp. DTU429	Sy05_BG			
Synergistales sp. DTU085	UC02_BG			
Syntrophomonadaceae sp. DTU125	Fi83_BG			
Syntrophomonadaceae sp. DTU366	Fi100_BG			
Syntrophomonadaceae sp. DTU375	Fi25_BG			
Syntrophomonadaceae sp. DTU378	Fi34_BG			
Syntrophomonadaceae sp. DTU385	Fi56_BG			
Syntrophomonas sp. DTU018	Fi02_BG			
Syntrophomonas sp. DTU019	Fi21_BG			
Tepidanaerobacter sp. DTU063	Fi22_BG			
Thermodesulfobiaceae sp. DTU406	Fi90_BG			
Thermotogaceae sp. DTU111	Th01_BG			
Thermotogaceae sp. DTU430	Th02_BG			
Treponema sp. DTU425	Sp04_BG			
Treponema sp. DTU426	Sp07_BG			

FIG 1 Characteristics of the recovered PGs. The columns designate extended PG names, short identifiers (IDs), pie graphs representing the completeness of the PGs, pie graphs representing their contamination level, and a color code indicating PGs that were firmly attached to the grass (green), planktonic (light blue), and those that were similarly enriched in both fractions (gray).

(Fig. 2). Some fundamental functions, such as methanogenesis, are preferentially restricted to the liquid fraction, as verified by *Methanomicrobiales* species (Eu01_BG, Eu02_BG, and Eu05_BG) that were identified only in the planktonic microbiome (Fig. 1 and 3). Other archaeal species belonging to the orders *Methanobacteriales* (Eu03_BG) and *Methanosarcinales* (Eu06_BG) have an "intermediate behavior," as they were identified both in FG and in the LG fractions. Relative abundance of bacteria and archaea in the two fractions was verified using real-time PCR with universal primers (27F/1492R and 109F/1492R). Results obtained revealed a 2.1-fold enrichment of bacteria in the

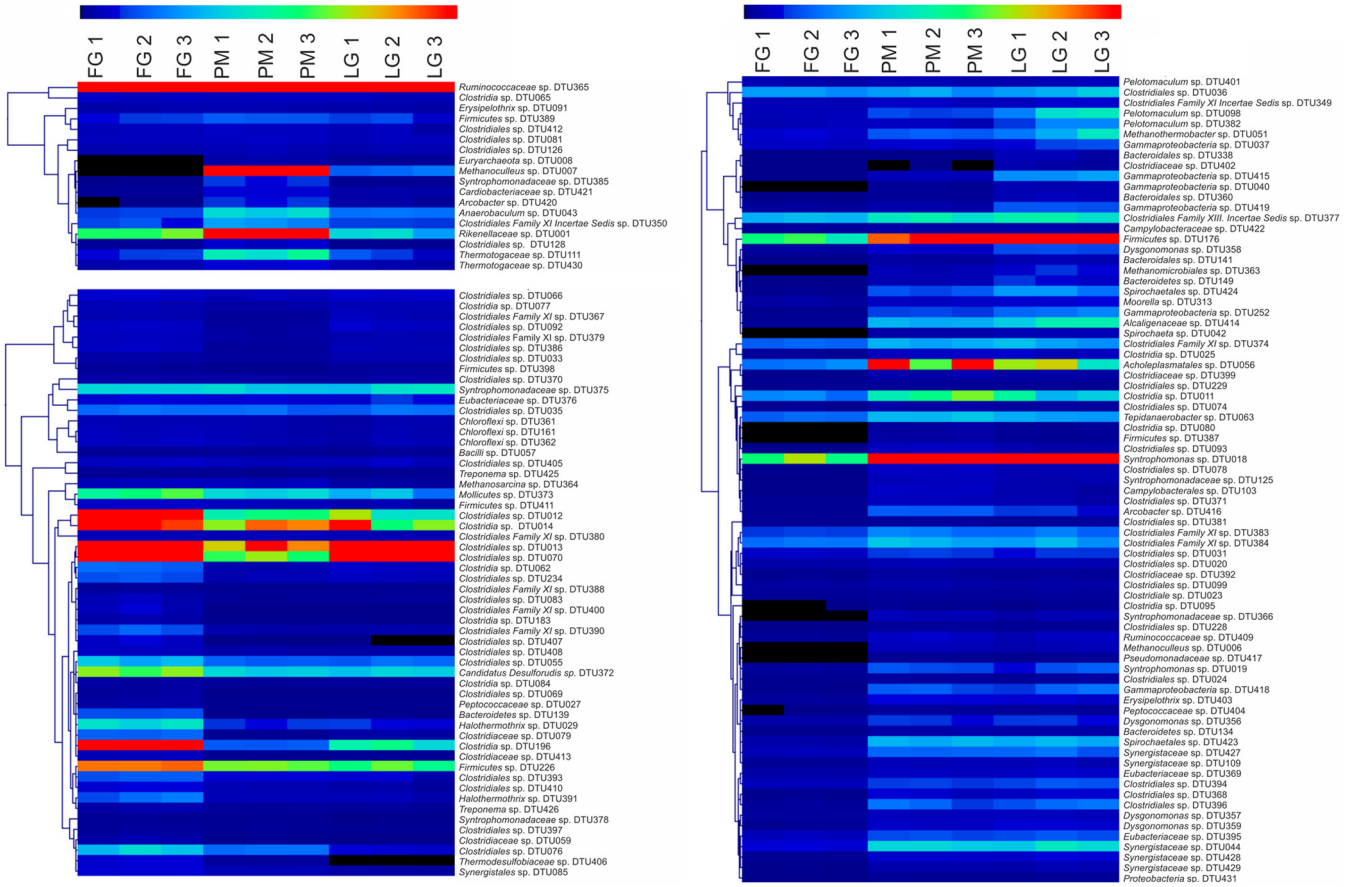

FIG 2 Abundance of the 151 PGs is represented as a heat map. Three replicates were collected from the microbial communities that were firmly attached to the grass (FG), residing in the liquid phase of the reactor (LG), or populating the reactor treating exclusively pig manure (PG). For better visualization of the outcomes, the heat map was split into three sections based on the hierarchical clustering of the extracted PGs. Correspondence between colors and coverage profile is illustrated with a scale (i.e., decreasing coverage is ranked from red to black) depicted above the heat maps.

firmly attached to grass fraction and a 4.8-fold enrichment of archaea in the liquid fraction.

A comparison between the gene contents of firmly attached to grass PGs and the planktonic PGs were the initial stepping stone to identifying the functional differences among these two groups. COG analysis showed that the number of genes in 9 clusters of orthologous groups of proteins was significantly different between the two microbial groups (Supplemental Data Set S1). Specifically, the firmly attached to grass PGs were more enriched in categories G (carbohydrate transport and metabolism), T (signal transduction mechanisms), and N (cell motility), while the planktonic PGs were more enriched in category C (energy production and conversion). The presence of numerous genes belonging to the planktonic PGs in COG category C (energy production and conversion) is related to their ability to use intermediate metabolites derived from plant polysaccharide degradation, such as volatile fatty acids, for energy production.

A deeper insight into the diverse metabolic properties of the firmly attached to grass and the planktonic PGs can be obtained from the SEED subsystem (second level annotation), which provides a more specific functional assignment of the proteins (Supplemental Data Set S1). This analysis confirmed that the firmly attached to grass PGs encode more proteins related to motility and chemotaxis, and additionally revealed an enrichment in adhesion and biofilm formation functional categories. The potential involvement of PGs in biofilm formation was also supported by the presence in these PGs of a higher number of proteins associated with the RNA polymerase sigma-54 factor gene *rpoN*, which is a pleiotropic transcription factor involved in regulation of

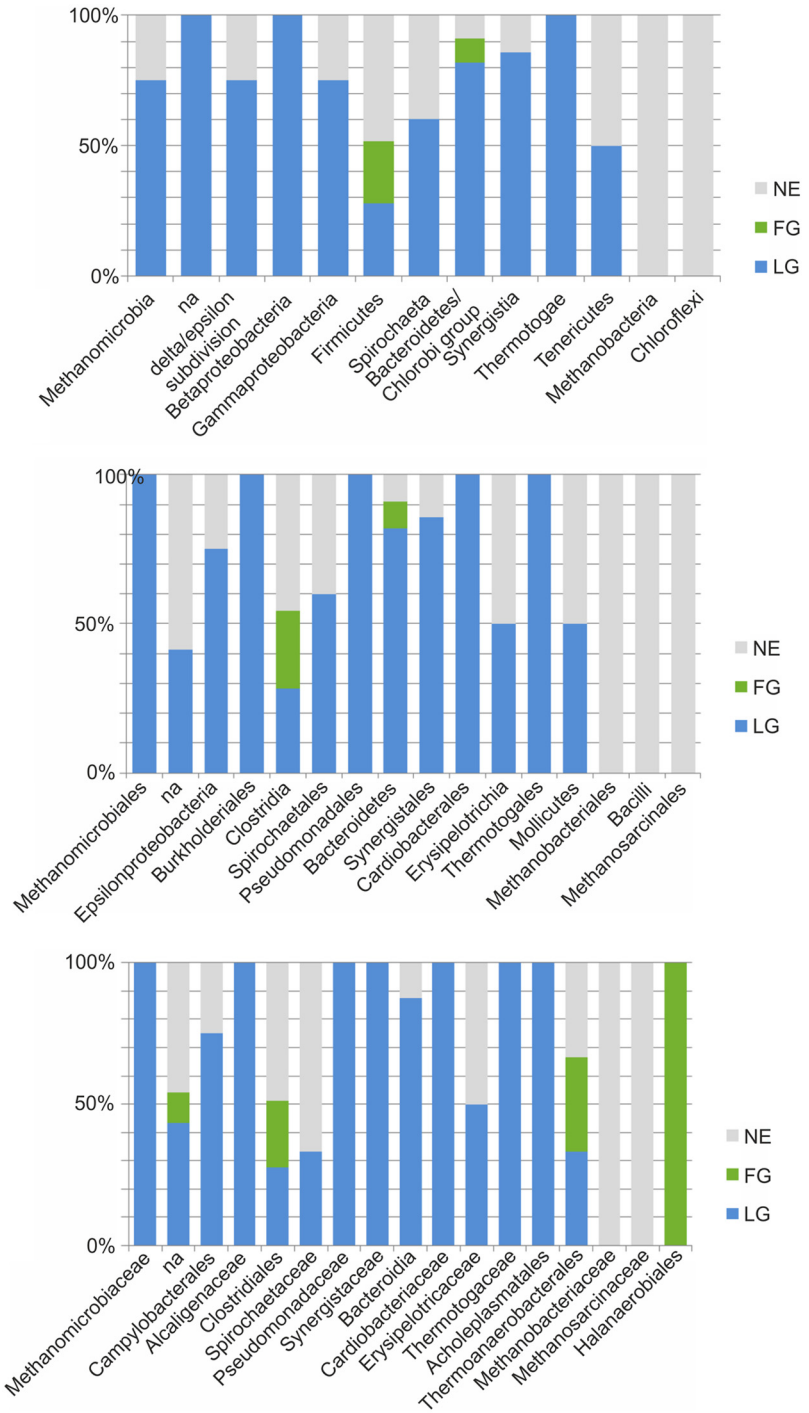

FIG 3 Comparative taxonomic analysis of the PGs. The PGs were grouped as firmly attached to the grass (green), planktonic (blue), and those that were similarly enriched in both fractions (gray). For each taxon, the proportions of PGs belonging to the three groups are represented. The PGs are classified at the phylum level (top), at the family level (middle), and at the order level (bottom).

genes related to carbohydrate metabolism, motility, and biofilm formation (24). Moreover, the ability of the firmly attached to grass PGs to perform carbohydrate degradation and metabolism is proven by the presence of numerous genes involved in disaccharide and oligosaccharide and/or monosaccharide utilization and by the presence of glycoside hydrolases. The planktonic PGs encode, on average, a higher number of proteins involved in transport and utilization of simple carbon compounds derived

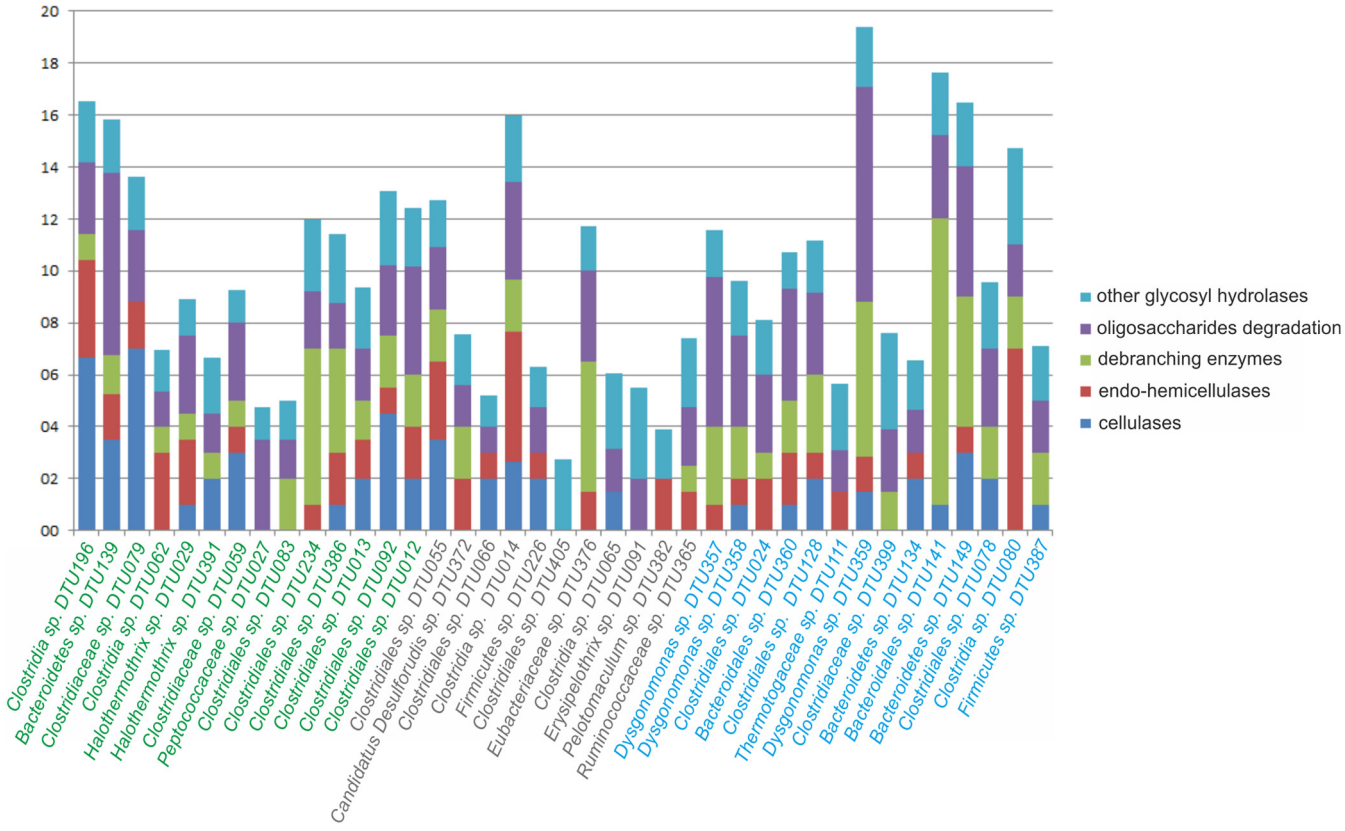

FIG 4 Number of carbohydrate-active domains in PGs. PGs with higher numbers of glycoside hydrolases (GH) domains and with estimated completeness higher than 50% were analyzed. Domains belonging to the GH were grouped into four classes based on their role in carbohydrate degradation, and the remaining GH are described as "other GH." PGs were colored depending on their spatial location, as follows: green, firmly attached; blue, liquid fraction; and gray, intermediate behavior.

from the degradation of complex plant polysaccharides, such as organic acids and tricarboxylate transporters. Moreover, they contain higher numbers of proteins involved in "one-carbon metabolism", "cofactors, vitamins, prosthetic groups, pigments", and "coenzyme F420" (groups defined by the SEED system), demonstrating a closer involvement during the process of methanogenesis.

The results from the dbCAN database (Supplemental Data Set S1) showed that the detected glycoside hydrolase included known carbohydrate-active enzymes, which were classified into five categories, namely, cellulases, endohemicellulases, debranching enzymes, oligosaccharide-degrading enzymes, and other glycoside hydrolases. The PGs having the highest numbers of glycoside hydrolase domains for each category are illustrated in Fig. 4. The analysis demonstrated that the firmly attached to grass PGs (particularly *Clostridia* sp. DTU196, *Clostridiaceae* sp. DTU079, and *Clostridiales* sp. DTU092) have more cellulases, while PGs in the liquid fraction (particularly *Bacteroidetes* sp. DTU149, *Dysgonomonas* sp. DTU359 and *Clostridiales* sp. DTU078) have more debranching enzymes.

Among the nearly complete population genomes, *Clostridiaceae* sp. DTU079, *Bacteroidetes* sp. DTU139, and *Clostridia* sp. DTU196 exhibited the highest enrichment in the FG samples (Supplemental Data Set S1). More specifically, the coverage profile of these PGs was found to be significantly increased in the FG samples compared to those of the LG samples by 14-fold ($P < 0.001$), 19-fold ($P < 0.0003$), and 25-fold ($P < 0.003$), respectively. Analysis of the carbohydrate binding modules among the reconstructed PGs evidenced that only the *Firmicutes* genomes of *Clostridiaceae* sp. DTU079 and *Clostridia* sp. DTU196 contain numerous cohesin and dockerin domains. The cohesin-dockerin pairs constitute fundamental elements of cellulosome, which in turn can

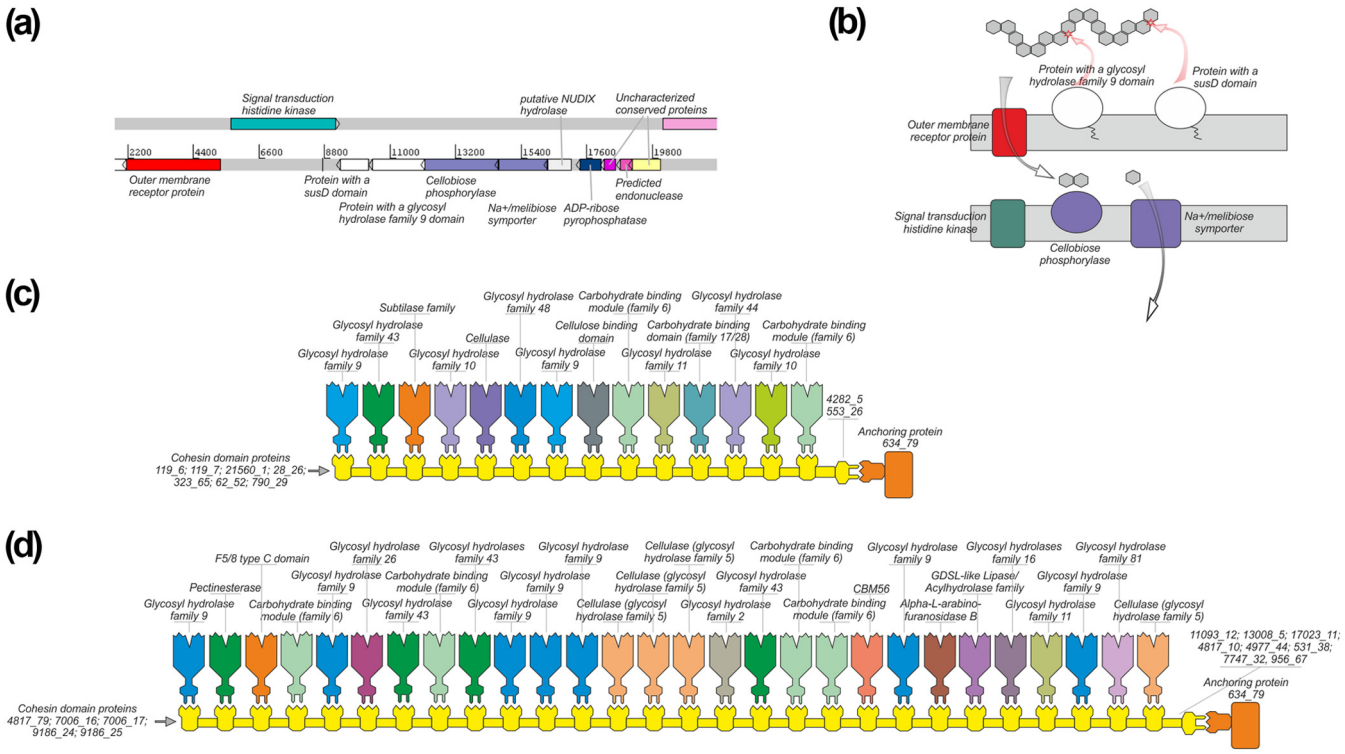

FIG 5 Polysaccharide utilization locus (PUL) organization and cellulosomes of Fi04_BG and Fi53_BG. (a) Gene organization of a relevant PUL in Ba06_BG. Genes are colored according to their COG functions. (b) Proteins encoded by the genes putatively involved in cellulose degradation are assigned to their cellular localization, and their putative role is reported according to the functional annotation and the model proposed by Martens and colleagues (35). (c and d) Schematic representation of the cellulosome complex identified in Fi04 and Fi53. The putative functional role is reported according to their annotation.

efficiently degrade plant polysaccharides (e.g., cellulose). The cellulosome structure of *Clostridiaceae* sp. DTU079 contains 12 cohesin and 61 dockerin domains (distributed on 6 and 36 proteins, respectively) and is much more complex in comparison to *Clostridia* sp. DTU196, which contains 12 cohesin and 26 dockerin domains (on 7 and 17 proteins, respectively) (Fig. 5; Supplemental Data Set S1). COG annotation of the *Clostridiaceae* sp. DTU079 cellulosome revealed that many of its enzymatic activities are represented by beta-xylosidases and endoglucanases, while the *Clostridia* sp. DTU196 cellulosome contains mainly beta-xylosidases. The evaluation of the genomic localization of genes encoding cellulosome components demonstrated that they are comprised in defined polysaccharide utilization loci. In *Clostridia* sp. DTU196, there are seven main regions, localized on scaffolds 13, 48, 286, 790, 944, 1201, and 1476, while in *Clostridiaceae* sp. DTU079, there are nine main regions on scaffolds 531, 956, 1421, 4817, 4977, 7006, 8003, 8935, and 9186. Moreover, some other genes encoding subunits of the cellulosome are scattered along the genome.

In contrast, the genes involved in the mechanism of polysaccharide utilization in *Bacteroidetes* sp. DTU139 are clustered in six specific regions (i.e., on scaffolds 3625, 9343, 18620, 21571, 22665, and 22746). The most interesting polysaccharide utilization loci was found on scaffold 18620, and it is composed of 8 genes, encoding, respectively, an outer membrane receptor protein (18620_3), a signal transduction histidine kinase (18620_4), two proteins of unknown function (18620_5, 18620_6), a cellobiose phosphorylase (18620_7), and a disaccharide transporter (18620_8) (Fig. 5a). The unknown protein 18620_5 contains a SusD (starch-binding protein) domain, while 18620_6 contains a glycosyl hydrolase family 9 domain.

DISCUSSION

The high completeness level of the recovered PGs allowed a detailed determination of the metabolic potential of the microbial population. The taxonomic assignment of

the microbial members showed that the dominant phyla of the microbiome were in accordance with those in previous studies (2, 9, 25). It should be underlined that 80 out of the 151 recovered PGs were newly identified, while the rest of the PGs (i.e., 71 PGs, 47% of the total community) were the same species found in previous metagenomic binning studies (see Fig. S1 in the supplemental material) (26, 27). This result clearly demonstrates that the engineered anaerobic digestion microcosm is far from elucidated, and further investigations are necessary to identify the numerous species that currently remain unknown at genome level. The newly identified PGs represented 53% of the planktonic microbiome and 40% of the community that was firmly attached to grass biomass. The relatively higher portion of newly identified PGs in the planktonic microbiome can be attributed to the utilization of pig manure as a cosubstrate. As extensively described in Supplemental File S1, the diverse chemical composition of pig manure (mainly due to high ammonia concentration) compared to that of cattle manure affected both the bacterial and archaeal community profile. The distinctively different microbial communities of biogas reactors treating cattle manure or pig manure are evidenced by plotting the coverage profiles of their metagenomes as a heat map (Supplemental File S1). It has been previously reported that environmental variables, such as the physicochemical conditions of the feedstock, govern microbial community structure and its linkage with process efficiency in anaerobic digesters (28).

The enrichment of specific microbes in the firmly attached to grass biomass could classify them as putatively responsible for lignocellulose degradation. More specifically, these microbes could bind to grass particles and fibers that are present in the reactor, and consequently, the proximity of the microorganisms with the substrate could facilitate the plant cell wall deconstruction process. It was shown that only a minor fraction of the microbiome (i.e., less than 16% of the microbial species) specialized in polysaccharide degradation, while all the remaining planktonic species were involved in different steps of the anaerobic digestion food chain.

All of the firmly attached to grass PGs belong to the *Firmicutes* and *Bacteroidetes*/*Chlorobi* phyla (Fig. 3), confirming previous findings based on 16S rRNA clone library analysis (17). The identification of a large number of *Firmicutes* species was expected, since it is one of the dominant phyla in proficient plant biomass-degrading ecosystems like the rumen (22). At a lower taxonomic level, these PGs were assigned to the orders *Clostridiales* and *Thermoanaerobacteriales*. In constrast, the identification of species belonging to the phylum *Bacteroidetes* in the firmly attached to grass sample is intriguing because uncultured species assigned to this taxonomic group have been recently identified in the rumen and also in other cellulose-degrading microbial communities (19, 21). Additionally, the numerical predominance of uncultured *Bacteroidetes* species in lignocellulose-degrading environments suggests a relevant contribution of these microbes to polysaccharide degradation (22, 29, 30).

The genomic properties of the extracted PGs were investigated to determine the rationale behind their spatial localization, emphasizing the microbes having a putative role in lignocellulose decomposition. The results from COG analysis showed that the firmly attached PGs were more enriched in functional categories T (signal transduction mechanisms) and N (cell motility). Despite the fact that COG functional categories are very general, this result indicates that once grass particles are introduced into the anaerobic system, specific microbial species are able to perceive this signal, move toward the plant material, and ultimately attach firmly onto it so as to efficiently metabolize and internalize carbohydrates. Previous studies demonstrated that chemotaxis and motility are widely diffused properties in lignocellulose-degrading systems; different polysaccharides, such as cellobiose, cellotriose, D-glucose, xylobiose, and D-xylose, can serve as chemoattractants for cellulolytic bacteria that migrate toward plant materials and decompose organic substances (31, 32). Moreover, the outcomes from the dbCAN database demonstrated that the PGs that were firmly attached to the grass particles had more genes encoding cellulases, while the planktonic PGs had more debranching enzymes. This finding suggests a cooperative function between the two microbial clusters. The exoglucanases of PGs present in the liquid fraction or in

Eubacteriaceae species (which were similarly enriched in liquid and grass fractions), such as *Eubacteriaceae* sp. DTU376, could prevent a potential accumulation of tetrasaccharides and cellobiose. In other cases, the accumulation of these compounds can in turn inhibit the cellulase activity of the firmly attached microbial species. In a previously studied thermophilic cellulose methanation consortium, a similar cooperative mechanism was found to be performed by *Spirochaetales* or *Thermotogales* species with *Clostridiales* species (19).

An interesting finding was related to the mechanisms that are followed by the different microbial species to bind and degrade lignocellulose. As reported in the Results, numerous cohesin-dockerin pairs were found in *Clostridiaceae* sp. DTU079 and *Clostridia* sp. DTU196. Cohesin modules function as the main building blocks of cellulosomal scaffoldin, while dockerin modules anchor the catalytic enzymes to the scaffoldin or can be present in the C termini of the same proteins (29, 33). The high efficiency of cellulosome in binding the bacterial cell to the plant polysaccharides explains why the PGs possessing this feature are highly enriched in the fraction that was firmly attached to the plant biomass.

In contrast, the mechanism of polysaccharide utilization in *Bacteroidetes* sp. DTU139 is completely different from that present in *Clostridiales* species. An initial distinct change is related to the absence of cohesin and dockerin domains in *Bacteroidetes* species. The enzymes catalyzing the conversion of cellulose in uncultured *Bacteroidetes* species are not arranged in a cellulosome (21), despite a previously reported cellulosomal scaffoldin gene in *Bacteroides cellulosolvens* (34). Considering the model of the starch utilization system (Sus) of *Bacteroides thetaiotamicron* (35), we can depict a polysaccharide utilization locus, where the cellulose is degraded to cellobiose by the glycosyl hydrolase family 9-containing enzymes, transported to the periplasmic space by the disaccharide transporter, cleaved to glucose by the cellobiose phosphorylase, and finally transported to the cytoplasm (Fig. 5b). It can be reasonably assumed that the outer membrane receptor protein and the outer membrane sensor are involved in the regulation of the other genes present in this locus.

This model extends the role of *Bacteroidetes* species in polysaccharide degradation beyond grass-degrading environments, such as the rumen, suggesting a more general involvement in anaerobic digestion systems. Our findings further show that important multifunctional enzymes, such as the cellulosome, can be remarkably different in species both due to the number and due to the roles of the subunits involved. In the present study, metagenomic binning integrated with bioinformatics strategies for gene annotation represents an important tool to identify and investigate microbial species participating in polysaccharide degradation. This approach proves to be particularly useful to explore the unculturable species present in the anaerobic digestion microbial community, which are impossible to investigate by classical genomic approaches. An accurate genome annotation elucidated the diverse metabolic strategies that are followed by specific microbial species to bind to and degrade plant material, providing to other community members the subproducts for methane production. Moreover, the genome sequencing and functional annotation of 25 PGs representing the microbial fraction firmly attached to the plant material is a fundamental step toward the comprehension of the mechanism for polysaccharide utilization in biogas reactors. In particular, the three draft genome sequences of anaerobic digesters that were adherent to the grass are themselves perfect candidates for further biotechnological applications in cellulose degradation, such as production of platform chemicals (e.g., 3-hydroxypropionic acid, xylitol, etc.) or sustainable alternative biofuels (e.g., biobutanol, aviation fuels, etc.).

MATERIALS AND METHODS

Biogas reactor configuration, DNA extraction, and metagenome sequencing. Samples were collected during the steady-state periods (i.e., stable biogas production with a daily variation of lower than 10% for at least 10 days) of two biogas reactors, as previously described (36). The first reactor was codigesting pig manure together with ensiled meadow grass, while the second reactor was treating only pig manure (monodigestion process). Three replicate samples were collected from each reactor; genomic DNA extraction and sequencing were performed independently for each replicate to allow statistical

evaluation of the population genome abundance in the different fractions. For the samples collected from the codigesting reactor, approximately 3 g of grass were separated from the liquid fraction. Plant fibers present in the liquid fraction (i.e., sample denoted "LG") were removed by driving the sample through a 40-μm nylon cell strainer filter and by centrifugation at 500 $\times$ *g* for 15 min. After this step, microbial cells were recovered with a centrifugation at 10,000 $\times$ *g* for 20 min at 4°C. The grass obtained after the separation of the liquid material was processed in order to recover the firmly attached microbes (i.e., sample denoted "FG"), as previously described (37). The grass was washed with basal anaerobic (BA) medium (38) to remove loosely attached microbial cells. After this step, the firmly attached microbes were stripped off by incubating the grass at 4°C in anaerobic conditions with BA medium containing 0.15% Tween 80 and, subsequently, by homogenizing the grass with a VDI-12 homogenizer (2 $\times$ 2 min bursts; VWR, Radnor, PA). The fragmented plant material remaining after homogenization was removed with a mild centrifugation step (500 $\times$ *g* for 15 min), and adherent cells remaining in the supernatant were recovered by centrifugation at 10,000 $\times$ *g* for 20 min at 4°C. Samples collected from the reactor performing the monodigestion (i.e., sample denoted "PM") were processed similarly to the liquid fraction collected from the codigestion process.

Genomic DNA from each sample was extracted using the PowerSoil DNA isolation kit protocol (Mo Bio Laboratories, Carlsbad, CA). The quality and concentration of extracted DNA were determined using NanoDrop (Thermo Fisher Scientific, Waltham, MA) and a Qubit fluorometer (Life Technologies, Carlsbad, CA), respectively. Integrity of the genomic DNA was evaluated with agarose gel electrophoresis. The sequencing was performed using the Nextera DNA library preparation kit (Illumina, San Diego, CA). Libraries were paired-end sequenced (2 $\times$ 150 bp) by a NextSeq 500 sequencer (Illumina, San Diego, CA) using the mid-output kit.

***De novo* metagenome assembly and binning process.** *De novo* assembly was performed with CLC Genomics workbench software v 5.1 (CLC bio, Aarhus, Denmark), using CLC's *de novo* assembly algorithm (kmer, 63; bubble size, 60; and minimum scaffold length, 500 bp). The binning process used to extract the population genomes from the assembly was previously reported by Campanaro et al. (26). Specific modifications and detailed parameters followed in the current study are provided in Supplemental File S1. It was previously demonstrated (39) that in the binning methods based on coabundance clustering, the specificity of the genome extraction process suddenly reaches the maximum level when more than 15 different samples are used. For this reason, additional samples from previous metagenomic studies were included in the analysis (12, 26, 27). The additional samples were only used to assist the alignment of the scaffolds and to increase the total number of coverage conditions examined. The completeness and the genome contamination of each PG were determined by considering the total number of essential genes and the number of duplicated genes in all the examined trials.

Taxonomic assignment of the population genomes. Taxonomic assignment of the PGs was performed with PhyloPhlAn software, using 400 broadly conserved proteins to extract phylogenetic signal (40). For 17 PGs whose assignment was at low or incomplete confidence with PhyloPhlAn, the result was verified and/or corrected using the Phylopythia classifier (41).

Identification of the PGs common between different metagenomic assemblies was performed by comparing the genomic sequences of the recovered PGs of the present study with those identified in previous metagenomic binning analyses related to biogas reactors and with the 60,585 genomes deposited in the NCBI Microbial Genome Resources database (accessed March 2016). The comparison was based on the calculation of the average nucleotide identity (ANI) value (42) of the protein-encoding nucleotide sequences, as previously described (27). Initially, a database with the nucleotide sequences of the PGs belonging to previous binning processes was obtained (26, 27). Similarity search was performed using BLASTn with all nucleotide sequences of the genes identified for each PG, using 1 $\times$ 10^{-5} as the minimum threshold. A script developed in-house, "ANI_calculator_CS.pl" (https://biogasmicrobiome.com/), determined the number of sequences in the BLASTN output having a match in the database and the ANI value for each PG. As previously defined, two genomes were considered to belong to the same species if at least 50% of the genes found a match and if the ANI was higher than 95% (42). Scaffolds encoding the 16S rRNA genes were identified using sequence similarity search of the Greengenes database, as previously proposed (43). Taxonomic assignment of the identified 16S rRNA genes was determined using RDP Classifier (44) with the confidence threshold set at 0.8.

Quantitative PCR. The real-time PCRs (RT-PCRs) were performed with the CFX96 real-time system (Bio-Rad, Hercules, CA). The PCR was performed using the iTaq SYBR green supermix (Bio-Rad, Hercules, CA) using the following conditions: preheating, 5 min at 95°C; and cycling, 40 cycles of 15 s at 95°C, 15 s at 50°C, and 60 s at 72°C. Due to the relevant size of amplicons obtained in the RT-PCR, the duration of the extension step was carefully optimized and the absence of nonspecific products was verified by checking the dissociation curve. Results were expressed in terms of fold difference between the two fractions examined (firmly attached to grass and liquid) as determined by the difference in the threshold cycle (C_T). Universal primers used for RT-PCR were 27F and 1492R for bacteria and 109F and 1492R for archaea.

Functional and statistical analysis. Gene finding was performed in the assembled scaffolds using Prodigal software set to metagenomic mode (45). All of the proteins predicted from the assembly were annotated using the reverse position-specific BLAST algorithm (RPSBLAST of NCBI BLAST+) using different databases, depending on annotation; for Clusters of Orthologous Groups of proteins (COG) the "COG only" RPSBLAST database was used (46), while for conserved protein domain prediction (Pfam) the Pfam RPSBLAST database was employed (47). BLAST results were filtered by E values lower than 1 $\times$ 10^{-5} and, additionally, only the best match was considered for COG. KEGG annotation was performed with the software GhostKOALA (48). Annotation of carbohydrate-active enzymes was performed using the software dbCAN (49). Functional

annotation was considered only for the PGs having an estimated completeness higher than 50%, as the high number of missing genes in incomplete PGs would affect the evaluation of the results. For KEGG, COG, and dbCAN annotation, the results obtained for all of the predicted proteins were separated and assigned to the single PGs using in-house developed Perl scripts. After the assignment process, the number of proteins belonging to each COG category (two-tailed *t* test, *P* value < 0.05), KEGG pathway, and carbohydrate-active module were calculated for each PG. The classification of the glycosyl hydrolases in different functional categories, such as cellulases, endohemicellulases, debranching enzymes, oligosaccharide-degrading enzymes, and other glycosyl hydrolases was obtained from Hollister and colleagues (50). The scaffolds of each PG were uploaded to the Rapid Annotations using Subsystems Technology (RAST) server and automatically reannotated using SEED and the RAST gene caller (51). Annotation results were downloaded, and the number of genes present in each first and second level SEED category was determined using in-house developed Perl scripts. A random sampling process was repeated 1,000 times with a Perl script implementing the Perl "rand()" function (52) to determine if the PGs are enriched in specific functional categories. The same numbers of proteins encoded in each PG were randomly collected from the entire assembly and assigned to each functional category. Finally, the fraction of random samples in which the number of genes assigned to a specific functional category was equal to or greater than *N* (where *N* is the number of genes assigned to the same functional category in the PG) was determined. The enrichment in a specific functional category was considered significant when the fraction was lower than the significance level ($\alpha = 0.05$).

Spatial assignment of the population genomes. The classification of the PGs as firmly attached to grass, planktonic, and with similar enrichment in both fractions (i.e., grass and liquid) was obtained by comparing the coverage of the PGs in the replicate samples, FG, and LG. Those enriched more than 2-fold in FG were considered firmly attached, those enriched more than 2-fold in the liquid fraction were considered planktonic, and the remaining PGs were assigned to the group with intermediate behavior. Consistency of the results among the three replicates was determined using a two-tailed *t* test and by selecting PGs with a *P* value lower than 0.05.

Accession number(s). Sequencing data described here can be found at the Sequence Read Archive under the identifier SRP074882 (BioProject accession number PRJNA319008).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.01244-18.

SUPPLEMENTAL FILE 1, PDF file, 0.7 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.9 MB.

ACKNOWLEDGMENTS

We acknowledge financial support from Energistyrelsen under the program EUDP, project 64013-0159, "New technology for an efficient utilization of meadow grass in biogas reactor."

Illumina sequencing was performed at the Ramaciotti Centre for Genomics (Sydney, Australia).

We declare no competing financial interests.

REFERENCES

1. Stewart EJ. 2012. Growing unculturable bacteria. J Bacteriol 194:4151–4160. https://doi.org/10.1128/JB.00345-12.
2. Schlüter A, Bekel T, Diaz NN, Dondrup M, Eichenlaub R, Gartemann KH, Krahn I, Krause L, Krömeke H, Kruse O, Mussgnug JH, Neuweger H, Niehaus K, Pühler A, Runte KJ, Szczepanowski R, Tauch A, Tilker A, Viehöver P, Goesmann A. 2008. The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. J Biotechnol 136:77–90. https://doi.org/10.1016/j.jbiotec.2008.05.008.
3. Wommack KE, Bhavsar J, Ravel J. 2008. Metagenomics: read length matters. Appl Environ Microbiol 74:1453–1463. https://doi.org/10.1128/AEM.02181-07.
4. Eikmeyer FG, Rademacher A, Hanreich A, Hennig M, Jaenicke S, Maus I, Wibberg D, Zakrzewski M, Pühler A, Klocke M, Schlüter A. 2013. Detailed analysis of metagenome datasets obtained from biogas-producing microbial communities residing in biogas reactors does not indicate the presence of putative pathogenic microorganisms. Biotechnol Biofuels 6:49. https://doi.org/10.1186/1754-6834-6-49.
5. Prosser JI. 2015. Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology. Nat Rev Microbiol Nature Res 13:439–446. https://doi.org/10.1038/nrmicro3468.
6. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.
7. Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. 2014. GroopM: an automated tool for the recovery of population genomes from related metagenomes. PeerJ PrePrints 2:e409v1. https://peerj.com/preprints/409/.
8. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
9. Treu L, Campanaro S, Kougias P, Sartori C, Bassani I, Angelidaki I. 2018. Hydrogen-fueled microbial pathways in biogas upgrading systems revealed by genome-centric metagenomics. Front Microbiol 9:1079. https://doi.org/10.3389/fmicb.2018.01079.
10. Vanwonterghem I, Jensen PD, Rabaey K, Tyson GW. 2016. Genome-centric resolution of microbial diversity, metabolism and interactions in anaerobic digestion. Environ Microbiol 18:3144–3158. https://doi.org/10.1111/1462-2920.13382.
11. Campanaro S, Treu L, Kougias PG, Luo G, Angelidaki I. 2018. Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. Water Res 140:123–134. https://doi.org/10.1016/j.watres.2018.04.043.
12. Kougias PG, Treu L, Campanaro S, Zhu X, Angelidaki I. 2016. Dynamic functional characterization and phylogenetic changes due to long chain

fatty acids pulses in biogas reactors. Sci Rep 6:28810. https://doi.org/10.1038/srep28810.
13. Treu L, Campanaro S, Kougias PG, Zhu X, Angelidaki I. 2016. Untangling the effect of fatty acid addition at species level revealed different transcriptional responses of the biogas microbial community members. Environ Sci Technol 50:6079–6090. https://doi.org/10.1021/acs.est.6b00296.
14. Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, Wilson C, Klingeman DM, Olson DG, Decker SR, Giannone RJ, Hettich RL, Brown SD, Lynd LR, Bayer EA, Himmel ME, Bomble YJ. 2016. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Sci Adv 2:e1501254. https://doi.org/10.1126/sciadv.1501254.
15. Shah AT, Favaro L, Alibardi L, Cagnin L, Sandon A, Cossu R, Casella S, Basaglia M. 2016. *Bacillus* sp. strains to produce bio-hydrogen from the organic fraction of municipal solid waste. Appl Energy 176:116–124. https://doi.org/10.1016/j.apenergy.2016.05.054.
16. Sträuber H, Schröder M, Kleinsteuber S. 2012. Metabolic and microbial community dynamics during the hydrolytic and acidogenic fermentation in a leach-bed process. Energy Sustain Soc 2:13. https://doi.org/10.1186/2192-0567-2-13.
17. Wang H, Vuorela M, Keränen AL, Lehtinen TM, Lensu A, Lehtomäki A, Rintala J. 2010. Development of microbial populations in the anaerobic hydrolysis of grass silage for methane production. FEMS Microbiol Ecol 72:496–506. https://doi.org/10.1111/j.1574-6941.2010.00850.x.
18. Cirne DG, Lehtomäki A, Björnsson L, Blackall LL. 2007. Hydrolysis and microbial community analyses in two-stage anaerobic digestion of energy crops. J Appl Microbiol 103:516–527. https://doi.org/10.1111/j.1365-2672.2006.03270.x.
19. Xia Y, Wang Y, Fang HHP, Jin T, Zhong H, Zhang T. 2014. Thermophilic microbial cellulose decomposition and methanogenesis pathways re-characterized by metatranscriptomic and metagenomic analysis. Sci Rep 4:6708. https://doi.org/10.1038/srep06708.
20. Smith SP, Bayer EA. 2013. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. Curr Opin Struct Biol 23:686–694. https://doi.org/10.1016/j.sbi.2013.09.002.
21. Naas AE, Mackenzie AK, Mravec J, Schückel J, Willats WGT, Eijsink VGH, Pope PB. 2014. Do rumen *Bacteroidetes* utilize an alternative mechanism for cellulose degradation? mBio 5:e01401-14. https://doi.org/10.1128/mBio.01401-14.
22. Edwards EJ, McEwan NR, Travis AJ, Wallace RJ. 2004. 16S library-based analysis of ruminal bacterial diversity. Antonie Van Leeuwenhoek 86:263–281. https://doi.org/10.1023/B:ANTO.0000047942.69033.24.
23. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
24. Hayrapetyan H, Tempelaars M, Nierop Groot M, Abee T, Kuipers O, Kovács A. 2015. *Bacillus cereus* ATCC 14579 RpoN (sigma 54) is a pleiotropic regulator of growth, carbohydrate metabolism, motility, biofilm formation and toxin production. PLoS One 10:e0134872. https://doi.org/10.1371/journal.pone.0134872.
25. Stolze Y, Bremges A, Rumming M, Henke C, Maus I, Pühler A, Sczyrba A, Schlüter A. 2016. Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. Biotechnol Biofuels 9:156. https://doi.org/10.1186/s13068-016-0565-3.
26. Campanaro S, Treu L, Kougias PG, De Francisci D, Valle G, Angelidaki I. 2016. Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnol Biofuels 9:26. https://doi.org/10.1186/s13068-016-0441-1.
27. Treu L, Kougias PG, Campanaro S, Bassani I, Angelidaki I. 2016. Deeper insight into the structure of the anaerobic digestion microbial community; the biogas microbiome database is expanded with 157 new genomes. Bioresour Technol 216:260–266. https://doi.org/10.1016/j.biortech.2016.05.081.
28. Ju F, Lau F, Zhang T. 2017. linking microbial community, environmental variables, and methanogenesis in anaerobic biogas digesters of chemically enhanced primary treatment sludge. Environ Sci Technol 51:3982–3992. https://doi.org/10.1021/acs.est.6b06344.
29. Flint HJ, Bayer EA, Rincon MT, Lamed R, White BA. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6:121–131. https://doi.org/10.1038/nrmicro1817.
30. Pope PB, Mackenzie AK, Gregor I, Smith W, Sundset MA, McHardy AC, Morrison M, Eijsink VGH. 2012. Metagenomics of the Svalbard reindeer rumen microbiome reveals abundance of polysaccharide utilization loci. PLoS One 7:e38571. https://doi.org/10.1371/journal.pone.0038571.
31. Monserrate E, Leschine SB, Canale-Parola E. 2001. *Clostridium hungatei* sp. nov., a mesophilic, N2-fixing cellulolytic bacterium isolated from soil. Int J Syst Evol Microbiol 51:123–132. https://doi.org/10.1099/00207713-51-1-123.
32. López-Mondéjar R, Zühlke D, Becher D, Riedel K, Baldrian P. 2016. Cellulose and hemicellulose decomposition by forest soil bacteria proceeds by the action of structurally variable enzymatic systems. Sci Rep 6:25279. https://doi.org/10.1038/srep25279.
33. Bayer EA, Belaich J-P, Shoham Y, Lamed R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554. https://doi.org/10.1146/annurev.micro.57.030502.091022.
34. Ding SY, Bayer EA, Steiner D, Shoham Y, Lamed R. 2000. A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. J Bacteriol 182:4915–4925. https://doi.org/10.1128/JB.182.17.4915-4925.2000.
35. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. 2009. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. J Biol Chem 284:24673–24677. https://doi.org/10.1074/jbc.R109.022848.
36. Tsapekos P, Kougias PG, Treu L, Campanaro S, Angelidaki I. 2017. Process performance and comparative metagenomic analysis during co-digestion of manure and lignocellulosic biomass for biogas production. Appl Energy 185:126–135. https://doi.org/10.1016/j.apenergy.2016.10.081.
37. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS, Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, Rubin EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331:463–467. https://doi.org/10.1126/science.1200387.
38. Bassani I, Kougias PG, Angelidaki I. 2016. *In-situ* biogas upgrading in thermophilic granular UASB reactor: key factors affecting the hydrogen mass transfer rate. Bioresour Technol 221:485–491. https://doi.org/10.1016/j.biortech.2016.09.083.
39. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto JM, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbeur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, Renault P, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol 32:822–828. https://doi.org/10.1038/nbt.2939.
40. Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304. https://doi.org/10.1038/ncomms3304.
41. Patil KR, Haider P, Pope PB, Turnbaugh PJ, Morrison M, Scheffer T, McHardy AC. 2011. Taxonomic metagenome sequence assignment with structured output models. Nat Methods 8:191–192. https://doi.org/10.1038/nmeth0311-191.
42. Rodríguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. Microbe 9:111–118.
43. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 31:533–538. https://doi.org/10.1038/nbt.2579.
44. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxon-

omy. Appl Environ Microbiol 73:5261–5267. https://doi.org/10.1128/AEM.00062-07.
45. Hyatt D, Locascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics 28:2223–2230. https://doi.org/10.1093/bioinformatics/bts429.
46. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261–D269. https://doi.org/10.1093/nar/gku1223.
47. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.
48. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
49. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. DbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–51. https://doi.org/10.1093/nar/gks479.
50. Hollister EB, Forrest AK, Wilkinson HH, Ebbole DJ, Tringe SG, Malfatti SA, Holtzapple MT, Gentry TJ. 2012. Mesophilic and themophilic conditions select for unique but highly parallel microbial communities to perform carboxylate platform biomass conversion. PLoS One 7:e39689. https://doi.org/10.1371/journal.pone.0039689.
51. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.
52. Treu L, Campanaro S, Nadai C, Toniolo C, Nardi T, Giacomini A, Valle G, Blondin B, Corich V. 2014. Oxidative stress response and nitrogen utilization are strongly variable in *Saccharomyces cerevisiae* wine strains with different fermentation performances. Appl Microbiol Biotechnol 98:4119–4135. https://doi.org/10.1007/s00253-014-5679-6.